

# Análisis de las Expresiones de la Violencia en Costa Rica: Tendencias, Factores Socioeconómicos y Respuesta Institucional en el Período 2020-2024

Natalia Abarca Madrigal, Yarling Henry Calderón, Nathalie Poveda Chaves, José Daniel  
Ramírez Aguilar, Christie Medina Leiton

*Ingeniería en Sistemas Computacionales, Universidad Fidélitas  
San Pedro, Costa Rica*

npoveda30053@ufide.ac.cr  
jramirez40724@ufide.ac.cr  
yhenry00049@ufide.ac.cr  
nabarca00271@iufide.ac.cr  
cmedina30255@ufide.ac.cr

**Resumen** – El estudio analiza expresiones de la violencia en Costa Rica mediante técnicas de Data Warehouse y herramientas de visualización, con datos del Poder Judicial y otras fuentes gubernamentales. Así se examinan patrones delictivos considerando variables socioeconómicas, demográficas y jurídicas, tomando en cuenta las tendencias en relación al nivel educativo, desigualdad económica o grupos etarios. Por medio de la identificación de estos factores permite diseñar políticas públicas basadas en datos.

**Palabras clave** - Análisis de datos, Criminalidad, Datasets, Data Warehouse, Estrategias preventivas, Femicidios, Violencia.

**Abstract** – The study analyzes expressions of violence in Costa Rica using Data Warehouse techniques and visualization tools, with data from the Poder Judicial and other government sources. Criminal patterns are examined considering socioeconomic, demographic, and legal variables, taking into account trends related to education level, economic inequality, and age groups. Identifying these factors enables the design of data-driven public policies.

**Keywords** - Criminality, Data analysis, Datasets, Data Warehouse, Femicides, Preventive Strategies, Violence.

## I. INTRODUCCIÓN

El presente trabajo tiene como fin detallar el proceso del análisis y modelado de datos sobre violencia en Costa Rica, haciendo uso de la información recopilada por el Poder Judicial y otros entes gubernamentales del país. Para ello, se utilizaron aplicaciones de Data Warehouse y análisis de datos como PowerBI y Tableau, de manera que se posibilite la identificación de patrones y se promueva el accionar informado de las instituciones públicas.

La inseguridad es un problema de múltiples aspectos que afecta a las sociedades y cuyo entendimiento demanda un estudio fundamentado en datos. Como es un fenómeno afectado por elementos sociales, económicos y políticos, su análisis debe tener en cuenta diversas dimensiones, incluyendo los datos accesibles en fuentes oficiales como los datasets del Poder Judicial.

Uno de los indicadores clave de la inseguridad es la cantidad de homicidios, cuyo incremento puede evaluarse a través de categorías legales particulares, lo que permite distinguir fenómenos como el tráfico de drogas, los femicidios y las compensaciones fiscales. Además, combinar estos datos con otras variables sociodemográficas enriquece la interpretación del problema. Por ejemplo, la correlación entre la violencia y los grados de educación puede desvelar patrones de incidencia delictiva, mientras que el estudio por quintiles de ingreso económico facilita la detección de posibles vínculos entre la desigualdad y la delincuencia. Del mismo modo, la investigación por edades facilita la evaluación del incremento de la violencia durante la adolescencia y la primera adultez.

La tecnología desempeña un rol crucial en las decisiones gubernamentales relacionadas con la seguridad. Mediante estrategias de análisis de datos, los diferentes poderes de la República pueden elaborar políticas públicas más eficaces y centradas en enfrentar la inseguridad desde un enfoque fundamentado en pruebas. El acceso a datos estructurados y pormenorizados simplifica la detección de patrones delictivos y la formulación de estrategias preventivas.

### A. Formulación del problema.

La inseguridad es un fenómeno complicado que se extiende a las sociedades actuales y está establecido por diversos factores, desde factores socioeconómicos hasta decisiones de políticas públicas. Por la extensa disponibilidad de datos y la expansión de las emergentes tecnologías, es posible examinar este problema con

un enfoque basado en datos, lo que facilita una mejor comprensión de sus dinámicas y sugiere soluciones más eficaces.

Uno de los componentes indispensables para el análisis de la inseguridad es la obtención y estudio de información acerca de homicidios y otros crímenes, que pueden identificarse de acuerdo a las categorías legales definidas por el Poder Judicial. Basados en estos datos, se pueden analizar fenómenos como el tráfico de drogas, los femicidios y los ajustes de cuentas, proporcionando una perspectiva más profunda de cómo y por qué se producen estos delitos.

También, el estudio de estos datos puede mejorar al cruzarlos con otras variables sociales. Por ejemplo, la relación entre los niveles de enseñanza y las expresiones de violencia favorece la evaluación de patrones en la incidencia de crímenes, mientras que el análisis de la exclusión social y los quintiles de ingreso económico puede analizar desigualdades estructurales que inciden en la delincuencia. Igualmente, el estudio por edad habilita entender el incremento de la violencia durante la adolescencia y la primera adultez, lo cual puede ser crucial para la construcción de estrategias preventivas.

En esta investigación, se analizará la forma en que la tecnología y el estudio de datos pueden ser recursos esenciales para la formación de políticas públicas dirigidas a disminuir la inseguridad. Mediante la utilización de estadísticas oficiales y bases de datos gubernamentales, el objetivo no solo será detallar el problema, sino también proporcionar una perspectiva fundamentada sobre posibles soluciones.

### **B. Objetivo General:**

Analizar factores de la inseguridad y expresiones de la violencia en Costa Rica desde un enfoque basado en datos obtenidos por el Poder Judicial, evaluando cómo la tecnología y el análisis de información pueden contribuir a la formulación de políticas públicas efectivas para su mitigación.

### **C. Objetivos Específicos:**

- Producir análisis de datos por medio de modelos para facilitar la identificación de patrones o tendencias relacionadas a la violencia en Costa Rica.
- Emplear herramientas de análisis de datos como Power BI y Tableau para una mayor eficiencia en la generación de reportes, simplificando el proceso de interpretación de múltiples fuentes de información sobre la violencia que puedan ser relevantes como el Poder Judicial de Costa Rica.
- Crear un data warehouse efectivo con la información obtenida por medio de los datasets de datos abiertos del Poder Judicial de Costa Rica, para así ordenarlos de manera íntegra y estructurada.
- Analizar la posible relación entre múltiples variables demográficas, socioeconómicas, políticas o jurídicas en patrones de violencia por medio del uso de herramientas de datos, para así facultar a las instituciones públicas sobre la situación actual del país en estas áreas.

### **D. Justificación**

La realización de la presente investigación surge como una necesidad y un apoyo indispensable para los entes públicos del país que se encargan de recopilar estos datos y trabajar día a día con ellos. En primer lugar, cualquier grupo de datos puede resultar

ineficaz sin realizar el adecuado análisis de los mismos. Un data warehouse, apoyado de modelos de datos y reportes efectivos permiten tomar decisiones basados en evidencia contundente. Es de esta manera en la que las distintas instituciones públicas, entes regulatorios y gubernamentales pueden adoptar medidas realmente diseñadas con total conocimiento del panorama actual. Todo esto puede traducirse en optimización de recursos institucionales, y robustecimiento de las redes de acción.

Por otro lado, como se mencionó anteriormente en la sección “Objetivos Específicos”, todo este proceso de construcción y modelado de reportes según estos datos permitirá identificar patrones o tendencias dentro de la información. De manera que se puedan predecir estos eventos o, al menos, determinar los factores de riesgo que pueden derivar en las mismas. Así, se puede tomar acciones que busquen ayudar activamente a los grupos, zonas o individuos en condición de vulnerabilidad. Es de esta manera que se optimizaría la labor de los entes gubernamentales encargados de informar y proteger a la población.

## **II. DESARROLLO**

### **A. Alcance**

En este proyecto se analizarán las expresiones de la violencia en Costa Rica durante el período 2020-2024. Se revisarán las tendencias que puedan llegar a influir en el incremento de estos delitos, entre ellos el factor socioeconómico por provincia a partir de los datos que almacena el SINIRUBE y los datos públicos del Poder Judicial. Adicionalmente, se realizará un análisis de la respuesta institucional a través de proyectos, instituciones, programas, políticas públicas y leyes existentes. El proyecto se enfoca en datos estadísticos nacionales, excluyendo casos de víctimas específicas publicadas en medios de comunicación.

Asimismo, para esto se espera responder a preguntas de caso definidas en este apartado con el fin de proponer desde el principio una metodología basada en cuestionamientos que permitan analizar de manera integral la temática, además se proponen preguntas que permitan realizar análisis adecuados contextualmente. Esto facilitará la elaboración de las diferentes visualizaciones de los datos. Algunas de esas preguntas son las siguientes.

- Tasa de violencia por provincia y cantón
  - ¿Cuáles son las zonas que tienen una mayor incidencia de delitos?
  - ¿Hay focos de violencia en determinadas áreas?
  - ¿Existen patrones entre provincias y su evolución en el tiempo?
- Análisis de violencia urbana vs. rural
  - ¿Dónde se concentran los delitos? ¿En zonas urbanas o rurales mayormente?
  - ¿Qué tipo de delitos se cometen en las zonas urbanas en comparación a los de zonas rurales?
- Tasa de violencia vs. índice de pobreza en el país
  - ¿Las zonas con mayor pobreza y desigualdad son las que registran más incidencia de criminalidad violenta?
  - ¿A lo largo del tiempo esta relación ha cambiado?
- Relación entre niveles educativos y tipos de delitos
  - ¿Cuáles son los grupos más expuestos a violencia según el grado educativo?
  - ¿Hay delitos en donde la correlación es más notoria?
- Análisis por grupos de edad
  - ¿En qué grupos etarios se reportan más víctimas de expresiones de la violencia?

- ¿Cuál es la diferencia etaria entre victimarios y víctimas?
- **Análisis por género de los delitos**
  - ¿Qué delitos afectan más a mujeres vs. hombres?
  - Comparación entre feminicidios y otros tipos de homicidio para evaluar el aumento respecto a otros crímenes.
- **Análisis de Violencia intrafamiliar vs. violencia pública**
  - ¿La violencia dentro del hogar tiene relación con el aumento de las expresiones de violencia?
  - ¿Hay algún patrón que indique que aumenta la violencia doméstica antes que otros tipos de violencia?
- **Violencia por estacionalidad**
  - ¿En qué meses del año hay mayor incidencia criminal?
  - ¿Los aumentos en las expresiones de la violencia en ciertos períodos pueden tener una correlación con la crisis social o económica?

## B. Limitaciones

El proyecto se enfrentará a que la calidad de los datos publicados pueda llegar a afectar la precisión del análisis, pueden existir reportes no archivados en las bases de datos y denuncias nunca realizadas. Además, algunos de los sitios webs consultados son repositorios secundarios, lo que puede generar problemas de recopilación y procesamiento de los datos. El proyecto no analizará variables como las psicológicas y/o culturales que puedan influir en las expresiones de la violencia.

## C. Requerimientos

Para llevar a cabo el análisis de datos en el proyecto, se determinaron los siguientes requerimientos:

### Funcionales:

- 1) Acceso y carga de datos
  - A. Identificación de fuentes de datos
    - i) Bases de datos oficiales: datos Abiertos del Poder Judicial, OIC, INEC y SINIRUBE.
    - ii) Formatos de datos: CSV, XLS, SQL, API
  - B. Extracción de los datos (ETL)
    - i) Extracción de los datos
    - ii) Transformación de los datos
      - (1) Limpieza de los datos (estandarización)
      - (2) Conversión de formatos y normalización
      - (3) Generación de claves primarias y relaciones entre tablas.
    - iii) Carga
      - (1) Inserción de datos en el Data Warehouse
      - (2) Organización en tablas y esquemas para consultas analíticas.
- 2) Procesamiento de datos y análisis estadístico
  - A. Estructuración del Data Warehouse
    - i) Diseño de modelo estrella o copo de nieve para consultas eficientes.
    - ii) Creación de dimensiones (por ejemplo: Tiempo, Ubicación, Tipo de Violencia, Características Socioeconómicas, etc.)
    - iii) Definición de hechos y métricas (por ejemplo, cantidad de delitos, tasas por región, índices de criminalidad)
  - B. Análisis de tendencias

- i) Identificación de picos de violencia en los diferentes períodos.
- ii) Comparación de datos entre períodos para detectar cambios.
- C. Análisis de correlaciones
  - i) Relación entre factores socioeconómicos, demográficos y las expresiones de violencia.
  - ii) Tasa de violencia por provincia y cantón
  - iii) Relación entre desempleo y criminalidad
  - iv) Comparación entre niveles educativos y tipos de delitos
  - v) Análisis por grupos de edad.
  - vi) Distribución por género de los delitos
  - vii) Violencia intrafamiliar vs. Violencia pública
  - viii) Violencia por la estacionalidad (meses, días, horas).
  - ix) Comparación entre años.
  - x) Presupuesto destinado a seguridad vs. tasa de criminalidad

## 3) Análisis de la respuesta institucional

- A. Evaluar cómo han reaccionado las instituciones a los cambios en los índices de violencia.
- B. Analizar tendencias y patrones en la violencia durante el período definido.

## 4) Generación de reportes y visualizaciones

- A. Dashboards interactivos
  - i) Creación de visualizaciones dinámicas que posean filtros que permitan categorizar la información por región, delito, período.
  - ii) Uso de gráficos de líneas, mapas de calor, histogramas, diagramas de dispersión, mapas de correlación, gráficos de áreas apiladas, gráfico de barras.
- B. Generación de informes
  - i) Reportes con análisis de tendencias y patrones.
  - ii) Exportación en diversos formatos.
- C. Comparación de datos en múltiples dimensiones
  - i) Análisis por provincia, cantón y distrito.
  - ii) Evaluación de diferentes patrones respondiendo a determinadas preguntas de caso.

### No funcionales:

- 1) Accesibilidad: se espera que la visualización y procesamiento de datos sea fácil de usar para analistas y funcionarios a quienes interese.
- 2) Escalabilidad: se espera que la visualización y procesamiento de datos permita la integración de más fuentes de datos para análisis futuros.
- 3) Disponibilidad: se espera que la visualización y procesamiento de los datos sean accesibles para consulta y generación de reportes.
- 4) Mantenimiento: se espera que se puedan ingresar nuevos datos o alterar la estructura del Data Warehouse.
- 5) Portabilidad: se espera que sea compatible con distintos sistemas operativos.
- 6) Fiabilidad: se espera que los datos sean precisos y verificables.
- 7) Usabilidad: se espera que la visualización y procesamiento de los datos faciliten el análisis e interpretación de los datos.

- 8) Compatibilidad: se espera que funcione correctamente dentro de las tecnologías establecidas como Power BI, Tableau y SQL Server.

#### D. Diccionario de Datos

Para la creación de la base de datos sin normalizar, primero fue necesario identificar los datasets y fuentes de información, en este caso son SINIRUBE y el OIJ. Así como un diccionario de datos donde poder ver las columnas de cada tabla, este diccionario de datos tiene una versión normalizada y una sin normalizar. El enlace de github donde se encuentran todos estos datos puede encontrarse en las referencias.

#### E. Diseño y Creación de la Base de Datos

Como parte de los primeros pasos necesarios para crear un datamart útil y personalizado para este proyecto, es necesario crear una base de datos que sobre la cuál montar el datamart. Es por ello que se definió la estructura de la Figura 1 como base de datos principal, considerando los datos históricos que conforman los datasets. En el siguiente diagrama se observan las tablas y sus respectivos atributos, así como sus llaves primarias, llaves foráneas y cómo se relacionan cada una de las tablas entre sí.

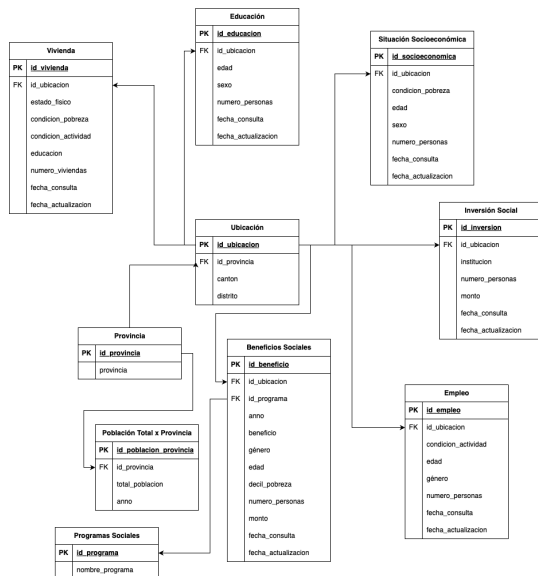


Fig. 1 Diagrama relacional de la base de datos propuesta para el presente proyecto.

Por otro lado, los scripts necesarios utilizados para crear la base de datos en SQL Server Management Studio pueden ser consultados en el repositorio destinado a este proyecto, en la dirección [https://github.com/chrisaiko/DataWarehouseProject\\_IQ\\_2025](https://github.com/chrisaiko/DataWarehouseProject_IQ_2025)

```

use master;

create database dump_dw_project
GO

use dump_dw_project
go

-- Situación socioeconómica por edad
CREATE TABLE situacion_socioeconomica_edad (
    Condicion_Pobreza VARCHAR(255),
    Edad INT,
    Provincia VARCHAR(255),
    Canton VARCHAR(255),
    Distrito VARCHAR(255),
    Numero_Personas INT,
    Fecha_Consulta DATE,
    Fecha_Actualizacion DATE
);

```

Fig. 2 Captura de pantalla de una parte del código perteneciente al archivo “script\_creación\_tablas\_sin\_normalizar.sql”.

De la misma manera, se creó un script con la normalización de la base de datos, ya que tal proceso es necesario para poder llevar a cabo todo el análisis de datos. Este archivo, denominado “Base de Datos normalizada.sql”, puede ser consultado en el repositorio de GitHub del proyecto anteriormente mencionado. Además, se ejemplifica también en la Figura 3.

```

USE dw_project_normalized;
GO

-- Dimensiones para Ubicación (Provincia, Cantón, Distrito)
CREATE TABLE dimension_provincia (
    id_provincia INT IDENTITY(1,1) PRIMARY KEY,
    nombre_provincia VARCHAR(255) UNIQUE
);

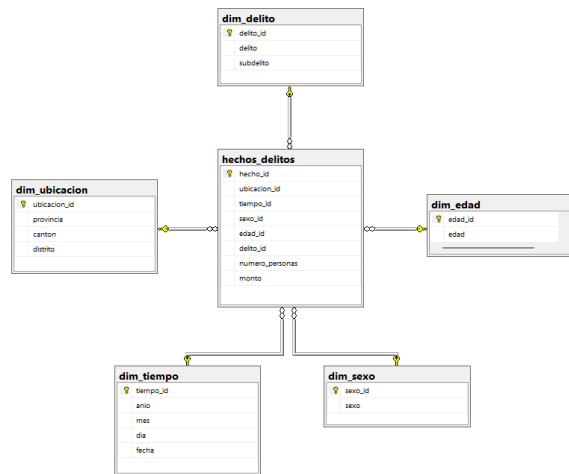
CREATE TABLE dimension_canton (
    id_canton INT IDENTITY(1,1) PRIMARY KEY,
    nombre_canton VARCHAR(255) UNIQUE,
    id_provincia INT,
    FOREIGN KEY (id_provincia) REFERENCES dimension_provincia(id_provincia)
);

CREATE TABLE dimension_distrito (
    id_distrito INT IDENTITY(1,1) PRIMARY KEY,
    nombre_distrito VARCHAR(255) UNIQUE,
    id_canton INT,
    FOREIGN KEY (id_canton) REFERENCES dimension_canton(id_canton)
);

```

Fig. 3 Captura de pantalla de una parte del código correspondiente al archivo “Base de Datos normalizada.sql”.

En otro orden de cosas, se construyó el código necesario para elaborar el datamart dentro del servidor y, de esta manera, poder realizar las visualizaciones de datos propias de esta herramienta. Al igual que en los casos anteriores, el script mencionado puede ser consultado en el repositorio del proyecto, bajo el nombre “Creación del Datamart.sql”. Debido a la naturaleza de este proyecto y de la composición de las tablas dentro de la base de datos, este datamart está construido en un modelo estrella. En la Figura 4 se observan las dimensiones propuestas; mientras estas convergen en la tabla de hechos denominada “Hechos\_Delitos”.



### III. REFERENCIAS

- [1] Poder Judicial de Costa Rica. (2022). Conjunto de datos de la Violencia Doméstica en formato XLS [Online]. Disponible: [https://ckanpj.azurewebsites.net/dataset/violencia-domestica/resource/ff1a0ba0-c676-4417-8ccd-aaa94ab2d3fc?view\\_id=471d53c6-e39e-4b2c-8ac3-71d36e156bc4](https://ckanpj.azurewebsites.net/dataset/violencia-domestica/resource/ff1a0ba0-c676-4417-8ccd-aaa94ab2d3fc?view_id=471d53c6-e39e-4b2c-8ac3-71d36e156bc4)
- [2] Poder Judicial de Costa Rica. (2022). Conjunto de datos de los Femicidios en formato CSV. [Online]. Disponible: [https://ckanpj.azurewebsites.net/dataset/femicidios/resource/30892570-8f29-43ba-8c55-b7a8b7b36781?inner\\_span=True](https://ckanpj.azurewebsites.net/dataset/femicidios/resource/30892570-8f29-43ba-8c55-b7a8b7b36781?inner_span=True)

[ENLACE DE GITHUB]

[https://github.com/chrisaqko/DataWarehouseProject\\_IQ\\_2025.git](https://github.com/chrisaqko/DataWarehouseProject_IQ_2025.git)